13

Clasificación digital de imágenes

Introducción

Una imagen de satélite es un conjunto de datos con información cuantitativa. Tal como se definió, la imagen digital es un modelo de reflectancia del terreno, que bien se puede referir a un solo intervalo espectral, en el caso de las imágenes pancromáticas monobanda, o bien puede estar constituida por un conjunto de bandas, si se trata imágenes multi o hiperespectrales. En cualquier caso, la información que proporciona la imagen está contenida exclusivamente en ese conjunto de datos. Ahora bien, la interpretación de tal información cuantitativa no es inmediata, especialmente para un usuario medio no familiarizado. Para este tipo de destinatario, es preferible etiquetar las áreas de la imagen con las características temáticas buscadas, para así facilitar su comprensión. Incluso el personal investigador puede requerir calificar la imagen para seleccionar áreas de interés o encontrar en la escena determinadas cualidades buscadas. La clasificación de imágenes se orienta en el campo de la teledetección a la ocupación de suelo o, si la resolución espectral de la imagen lo permite, a establecer categorías de las relacionadas en una leyenda previamente confeccionada.

La clasificación digital es un proceso de generalización temática que categoriza la información de la imagen, esto es, convierte la información cuantitativa en información cualitativa. Como en cualquier proceso de generalización cartográfica, siempre se pierde información con relación al conjunto original de datos, pero en el caso de la clasificación, esa merma está justificada por la mejora en la interpretación y en la usabilidad de la imagen.

El problema en el ámbito de la clasificación de imágenes espaciales queda planteado del siguiente modo: se trata de establecer a qué clase pertenece una determinada celda, de entre las i clases que integran la leyenda, en virtud de las componentes de su vector de características. Por ejemplo, se pretende establecer si una celda que presenta unos determinados niveles digitales en cada una de las distintas bandas de la imagen está ocupada por un suelo desnudo, un cultivo herbáceo, una zona urbana o un bosque de coníferas.

La imagen como variable aleatoria multivariante

En análisis digital de imágenes es muy común recurrir a métodos paramétricos de clasificación por la sencillez del tratamiento.

Los niveles digitales en una imagen de satélite pueden considerarse como una variable aleatoria. Será univariante cuando la imagen tenga una sola banda y multivariante en el caso de las imágenes multi o hiperespectrales. En la mayoría de los casos, la distribución de los ND tiende a la distribución normal. Solamente cuando en la imagen existen zonas bien con-

trastadas y diferenciadas, la distribución puede partirse, aunque cada una de las zonas responde a su vez a la forma gaussiana de distribución.

Considerar ajustada a la distribución normal el conjunto de ND de la imagen tiene la ventaja de poder representar la distribución solamente con dos parámetros: su media y su varianza en el caso de imágenes monobanda y su vector de medias y la matriz de covarianzas, en el caso de multiespectrales.

La función de densidad de la distribución normal univariante es:

$$f(z) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{z-\mu}{\sigma}\right)^2} \qquad (1)$$

y, de forma análoga, la de la normal multivariante es:

$$f(\vec{z}) = \frac{1}{\sqrt{|\Sigma|}\left(\sqrt{2\pi}\right)^n} e^{-\frac{1}{2}\left(\vec{z}-\vec{M}\right)^t \Sigma^{-1}\left(\vec{z}-\vec{M}\right)} \qquad (2)$$

donde $\vec{z}$ es el vector de características de una celda determinada, $\vec{M}$ es el vector de medias y Σ es la matriz de covarianzas, ya utilizada en la transformación a componentes principales, explicada en el capítulo 12:

$$\vec{z} = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{pmatrix}; \quad \vec{M} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{pmatrix}; \quad \Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_n^2 \end{pmatrix}$$

Considerando solamente dos bandas, la distribución de ND adquiere la forma de una nube bidimensional, con una mayor densidad de puntos en su zona central, en torno a lo que se podría denominar su centro de gravedad y que muestra habitualmente una cierta alineación o tendencia en una determinada dirección.

Si la imagen tiene tres bandas, en condiciones ideales, la nube de puntos tendrá una forma que podría asemejarse a un elipsoide. Si, finalmente, contiene cuatro o más bandas, la nube podría considerarse un hiperelipsoide. En cualquiera de los casos, para distribuciones normales, el agrupamiento de puntos en el espacio de características queda definido por su vector de medias, que señala el centro de gravedad de la nube, y la matriz de covarianzas de la imagen, que determina su forma. El hiperelipsoide puede obtenerse haciendo constante la forma cuadrática del exponente de la función de densidad multivariante:

$$\left(\vec{z}-\vec{M}\right)^t \Sigma^{-1}\left(\vec{z}-\vec{M}\right) = K \qquad (3)$$

que es el cuadrado de la llamada distancia de Mahalanobis entre $\vec{z}$ y $\vec{M}$ en la geometría de Σ.

La orientación de los ejes está definida por los vectores propios de la matriz de covarianzas, y el grado de dispersión o alargamiento de la nube en cada dirección depende del valor propio asociado a cada vector propio.

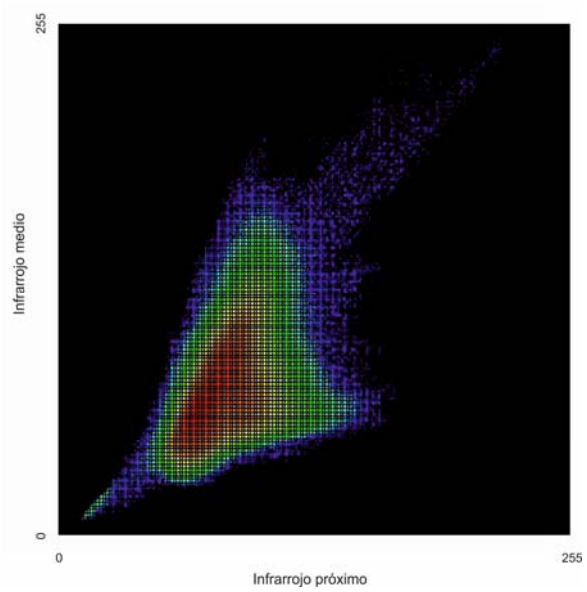
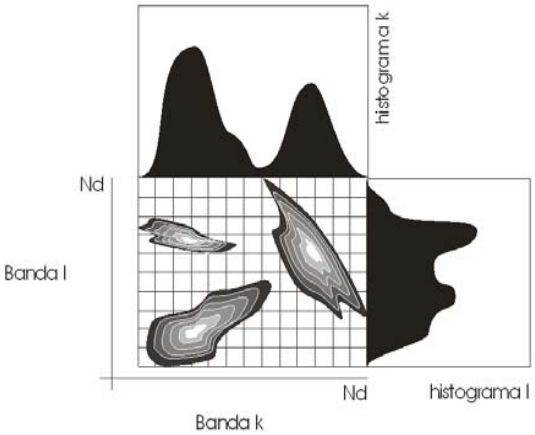


Imagen en color compuesto RGB=IRm-IRp-R y un diagrama de dispersión IRm/IRp con resolución radiométrica de 8 bits. Los colores cálidos indican mayor concentración de celdas.

Entrenamiento del clasificador

Es un hecho que grupos de celdas con un comportamiento espectral similar aparecerán en lugares próximos cuando se representan en un diagrama de dispersión. En el terreno, esas celdas tienen una misma naturaleza química, y, en definitiva, están ocupadas por la misma cobertura. En el proceso de identificar y discriminar categorías nominales en la imagen, ese conjunto de celdas constituirán una clase.



Tres agrupamientos de puntos pueden ser confundidos con dos si solamente se estudia la distribución de ND en una sola banda. El aumento de la espectralidad beneficia la identificación.

Significa esto que en el espacio espectral, las celdas se organizan como nubes de puntos próximos entre ellos y, a su vez, diferenciados de otras categorías de celdas. En una situación ideal, las nubes de puntos son independientes, alejadas unas de otras y claramente identificables y separables. En las imágenes reales, las nubes se mezclan, se solapan parcialmente y su discriminación no resulta tan fácil. Sin embargo, agrupamientos o racimos (*clusters*) de puntos que pueden ser confundidos como uno solo cuando se los estudia en un espacio espectral de tan solo unas cuantas bandas, pueden ser separables si se recurre a un mayor número de canales. La *espectralidad* de la imagen es un factor esencial a la hora de discriminar categorías.

Entendiendo *clasificador* como el algoritmo o el operador matemático de clasificación, el entrenamiento del clasificador es una tarea mediante la cual se definen los parámetros esenciales que identifican a cada clase, esto es, su vector de medias y su matriz de covarianzas.

El entrenamiento del clasificador puede ser de dos tipos: supervisado o no supervisado. En el primero, el operador delimita una muestra de celdas por cada clase, a partir de la cual se calculan ambos estadísticos. En el entrenamiento no supervisado, a partir de unas instrucciones mínimas, se deja al sistema de tratamiento que busque agrupamientos *naturales* de valores en el espacio de características.

Entrenamiento supervisado

Un paso previo al entrenamiento supervisado es la elaboración de la leyenda de la clasificación. El analista deberá decidir qué clases buscará en la imagen, y una vez establecidas las categorías, habrá de definir en la escena los campos de entrenamiento (*training fields*), que son un conjunto de parcelas constituidas por celdas suficientemente representativas de cada clase.

De la correcta selección de los campos de entrenamiento depende estrechamente el resultado de la clasificación. En primer lugar, la pureza de los píxeles es un requisito necesario. Las celdas recolectadas en los campos de entrenamiento deben representar solamente a la cubierta que se pretende caracterizar, sin contaminación de otras. Dicho de otro modo, hay que descartar los píxeles mixtos, situados sobre dos clases diferentes o en las proximidades de bordes de parcela con las categorías que se pretenden representar.

Pero a su vez, para que le muestra sea representativa, hay que respetar y recoger en ella la variabilidad interna de la clase. El conjunto de campos de entrenamiento debe incluir la variación interna de cada clase para dar cabida a las posibles condiciones diferencia-

das de la cubierta, como su orientación, la iluminación, el tipo de suelo, su grado de humedad, amén de las diferencias inherentes a la propia cubierta.

En cuanto al número y distribución de los campos de entrenamiento, utilizando uno solo para representar cada clase se corre el riesgo de particularizar excesivamente sus características. Así, tras la clasificación podrían aparecer zonas que, aun estando ocupadas por la misma cubierta, quedarían sin atribuir a esa clase. Por ejemplo, a la hora de entrenar un cultivo como el olivar, conviene tomar varios campos de entrenamiento en los que se incluyan celdas ocupadas por distintos tipos y condiciones de cultivo: tradicionales, intensivos, en secano, en regadío, sobre suelos rojos, sobre albarizas, en ladera norte y sur, con poca y mucha pendiente, etc.

El número de celdas necesario para configurar la clase en el entrenamiento depende de la variabilidad interna de la clase. Considerando que los ND de cada clase siguieran una distribución normal, el número mínimo de celdas a tomar para estimar un ND promedio representativo es:

$$n \geq \frac{z_{\alpha/2}^2 \sigma^2}{\varepsilon^2} \qquad (4)$$

donde $z_{\alpha/2}^2$ es el percentil de la distribución normal $N(0,1)$ correspondiente a la mitad del nivel de confianza exigido en la estimación, σ^2 es la varianza de la clase en la banda más desfavorable y ε es el error que se está dispuesto a admitir en la estimación del ND promedio.

Por ejemplo, para un intervalo de confianza del 95 % ($z_{0,05/2}$=1,96), una desviación típica de los ND de 15 unidades y un error máximo de 2 niveles digitales para la estimación del ND promedio, se necesita una muestra de al menos $n \geq 1,96^2 \cdot 15^2 / 2^2 = 216$ celdas. Para un nivel de confianza del 99 %, en cambio, se necesitarían $n \geq 371$ celdas.

En conclusión, clases muy uniformes quedarán bien descritas con pocas celdas. Por el contrario, clases de reflectancia muy heterogénea necesitarán un mayor número de celdas para estar bien caracterizadas.

Las celdas de entrenamiento deben estar distribuidas en campos o polígonos de entrenamiento, siendo preferible un mayor número de ellos, aunque sean de menor tamaño, que un sólo campo de entrenamiento que contenga todas las celdas de la muestra. En cuanto a su situación, los campos de entrenamiento deben distribuirse espacialmente según la variabilidad de la escena.

Cuando no se disponga de otra información, como aproximación se pueden tomar entre 10 y 100 celdas por banda de la imagen multiespectral para cada una de las clases. Así pues, para una imagen de 5 bandas, deberían recolectarse entre 50 y 500 celdas para configurar la muestra de entrenamiento de cada clase, distribuidas en varios polígonos.

Entrenamiento no supervisado

Los métodos no supervisados o automáticos buscan en el espacio de características de la imagen agrupaciones *espontáneas* de los datos, de acuerdo a alguna función de similaridad. En la literatura, esas nubes de puntos aparecen con diferentes denominaciones: agrupamientos, *clusters,* grupos o racimos, y pueden considerarse como regiones continuas del espacio de características que contienen densidades relativamente altas de puntos, y que se encuentran a su vez

separada de otras regiones de alta densidad por regiones cuya densidad de puntos es relativamente baja.

El entrenamiento no supervisado ordinariamente va unido al propio proceso de clasificación y recorre un triple proceso. En primer lugar, se determinan las características de las celdas de entrenamiento, a continuación, se calcula la similaridad o distancia estadística entre casos y, finalmente, se aplica algún criterio de agrupamiento de los casos similares o de escisión en clases diferentes.

La clasificación automática tiene el inconveniente de poder encontrar agrupamientos que realmente no tienen significado identificable en la leyenda de la clasificación. En realidad, desde un punto de vista teórico, siempre que no existan dos vectores de características exactamente iguales, es posible identificar tantas clases diferentes como celdas tenga la imagen, lo cual carece de sentido cartográfico. Por esta razón hay que informar al clasificador de numerosos detalles para que busque solamente los *clusters* que se necesitan, o un número próximo al deseado. En definitiva, el número de parámetros requeridos para que el algoritmo proporcione resultados consistentes, en numerosas veces desconocidos, hace dudar del propio apelativo de *no supervisado*.

Por el contrario, la clasificación supervisada puede adolecer del defecto contrario, esto es, disponer de una leyenda excesivamente ambiciosa que incorpora clases que espectralmente no son separables.

De todo esto se concluye que los métodos mixtos de clasificación, que incluyen a los dos anteriores, son deseables sobre cualquiera de ambos por separado, siempre que se disponga de los recursos necesarios para llevarlos a cabo. En la estrategia para combinar los procedimientos supervisados y no supervisados pueden seguirse dos caminos: se puede realizar una clasificación inicial no supervisada cuyo resultado será posteriormente supervisado por el analista, agrupando *clusters* de igual significación temática y separando en nuevas clases las que mezclen dos categorías diferentes, o bien se pueden hacer las dos clasificaciones, supervisada y no supervisada, y posteriormente se sintetizan los resultados obtenidos. Para las clases obtenidas pueden darse tres casos:

- **Categorías que aparezcan en las dos.** La leyenda es correcta porque las clases cartográficas propuestas en el entrenamiento supervisado tienen además significación espectral.
- **Categorías recogidas en la leyenda, pero no identificables en la clasificación no supervisada.** En este caso se ha sido demasiado ambicioso en el diseño de las clases y la espectralidad de la imagen no permite su separación. Por lo tanto, deben reestudiarse estratificando la imagen o, simplemente, simplificando la leyenda.
- **clases no supervisadas que aparecen sin estar en la leyenda.** La clasificación proporciona mayor número de clases espectrales que cartográficas y por ello el entrenamiento ha sido tímido. La información contenida en la imagen permite desagregar más la leyenda, adaptándola a la heterogeneidad de la escena.

Análisis de separabilidad

Con el fin de dimensionar adecuadamente la campaña de entrenamiento y tratar de conseguir que las clases cartográficas coincidan con las clases espectrales, es aconsejable realizar previamente un análisis de separabilidad de las clases. Esta información permitirá al analista seleccionar solamente las clases que el clasificador sea capaz de resolver como diferenciadas.

Si las clases son patentemente diferentes, el riesgo de error en la clasificación será bajo en general, pero probablemente la leyenda será también más pobre de lo que permitiría la riqueza espectral de la imagen. Por el contrario, si clases son muy similares desde el punto de vista espectral, la clasificación entrañará un elevado riesgo de confusión que requeriría, en su caso, sustituir la leyenda por otra menos desagregada. Antes de la fase de clasificación conviene, pues, estudiar la separabilidad de las clases.

Existen dos grupos de métodos para ello:

- **Métodos gráficos**, que evalúan la separabilidad a partir de los diagramas de signaturas.
- **Métodos estadísticos**, que evalúan la separabilidad de dos agrupamientos de datos midiendo la distancia entre ellos en el espacio de características.

Los primeros proporcionan una información preliminar para decidir *a priori* si las categorías serán o no discriminables. Un diagrama de signaturas es una generalización de la signatura espectral de las clases, sobre la que se remuestrea el barrido espectral de acuerdo con la anchura de banda y la longitud de onda central de los canales del sensor. Las gráficas incluyen el valor medio por clase en cada banda y alguna medida de dispersión alrededor de ella, como el rango de variación, la desviación estándar, etc. Para cada banda es posible estudiar más detalladamente el solapamiento estadístico de las clases superponiendo sus histogramas.

Los métodos estadísticos miden la distancia existente entre los centros de dos agrupamientos, es decir, la que hay entre sus vectores de medias. La distancia de Minkowski se define así:

$$D_p\left(\mathrm{M}_A,\mathrm{M}_B\right)=\left[\sum_{i=1}^{n}\left|a_i-b_i\right|^p\right]^{\frac{1}{p}} \qquad (5)$$

donde $p\geq 1$, y a y b son los componentes de los vectores de medias de los agrupamientos A y B. Dando valores a p se obtienen las distintas distancias métricas, de entre las cuales la más familiar es la distancia euclídea para n bandas, que se obtiene cuando p=2:

$$D_2\left(\mathrm{M}_A,\mathrm{M}_B\right)=\sqrt{\sum_{i=1}^{n}\left(a_i-b_i\right)^2} \qquad (6)$$

Escrito de otra forma, la distancia euclídea entre los centros de dos agrupamientos es:

$$\delta_{AB}=\sqrt{\sum_{i=1}^{n}\left(\overline{ND}_{A_i}-\overline{ND}_{B_i}\right)^2} \qquad (7)$$

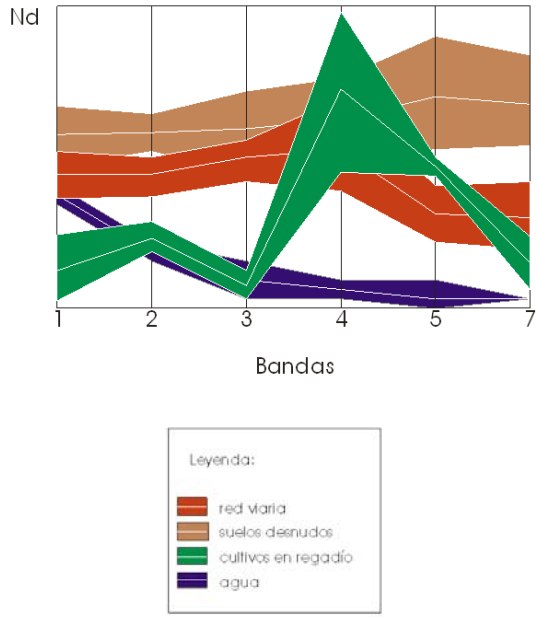


Diagrama de signatura de cuatro clases, para un sensor de 7 canales.

siendo $\overline{ND}_{A_i}$ y $\overline{ND}_{B_i}$ los niveles digitales medios en cada banda *i* de los agrupamientos *A* y *B*.

Otra forma de medir la distinguibilidad de las clases es mediante la separabilidad estadística, que se define así:

$$d_{AB} = \frac{|\mu_A - \mu_B|}{\sigma_A + \sigma_B} \tag{8}$$

Lo cual es un método similar al anterior, pero proporciona valores normalizados al dividir por la suma de las varianzas de ambos agrupamientos, es decir, relativiza las distancias de acuerdo a la dispersión de los valores.

En distribuciones normales multivariantes puede adoptarse la divergencia estadística como parámetro de separabilidad:

$$\begin{aligned} D_{AB} = {} & \tfrac{1}{2} Tr\left[\left(\Sigma_A - \Sigma_B\right)\left(\Sigma_B^{-1} - \Sigma_A^{-1}\right)\right] + \\ & + \tfrac{1}{2} Tr\left[\left(\vec{M}_A - \vec{M}_B\right)^t \left(\Sigma_A^{-1} + \Sigma_B^{-1}\right)\left(\vec{M}_A - \vec{M}_B\right)\right] \end{aligned} \tag{9}$$

Donde *Tr* indica la traza de las matrices resultantes.

Es también relativamente frecuente el uso de la divergencia transformada:

$$DT_{AB} = 2\left(1 - \tfrac{1}{8} e^{-D_{AB}}\right) \tag{10}$$

donde D_{AB} es la divergencia estadística entre *A* y *B* y siempre $0 \le DT_{AB} \le 2$. El criterio de interpretación es el siguiente:

- Si $DT_{AB} \ge 1{,}9$, las clases *A* y *B* son separables.
- Si $1{,}7 < DT_{AB} < 1{,}9$, la separación entre *A* y *B* es aceptable.
- Si $DT_{AB} < 1{,}7$, las clases *A* y *B* son difícilmente separables.

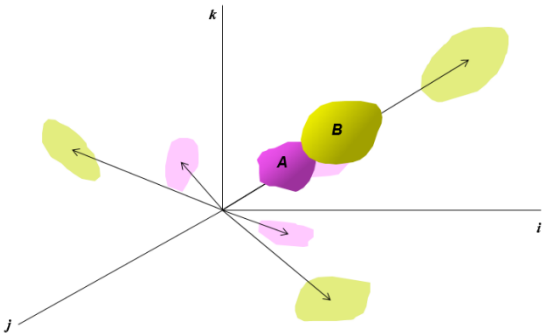
Un procedimiento muy utilizado para evaluar distancias entre clases en teledetección hiperespectral es medir el coseno del ángulo espectral, o el ángulo directamente, que da lugar a algoritmos como el mapeador de ángulo espectral (SAM). Dado que el producto escalar de los vectores de medias de dos agrupamientos *A* y *B* es:

$$\vec{M}_A^t \cdot \vec{M}_B = \left|\vec{M}_A\right| \cdot \left|\vec{M}_B\right| \cdot \cos\alpha$$

se puede escribir:

$$\alpha = \cos^{-1}\left[\frac{\vec{M}_A^t \vec{M}_B}{\left|\vec{M}_A\right|\left|\vec{M}_B\right|}\right] = \cos^{-1}\left[\frac{\sum_{i=1}^{n} a_i b_i}{\sqrt{\sum_{i=1}^{n} a_i^2}\sqrt{\sum_{i=1}^{n} b_i^2}}\right] \tag{11}$$

aunque a primera vista pueda parecer que agrupamientos diferentes que estén alineados con el origen de coordenadas espectrales serían confundidos con un solo agrupamiento según este criterio, puesto que $\cos\alpha = 1$, esta alineación habitualmente no ocurre en todas las bandas, por lo cual siempre se encontrará un cierto ángulo espectral distinto de cero. Solamente en el caso de que el alineamiento se verifique en el conjunto de todas las bandas, es cuando habrá que considerar que la clase es única, aun mostrando distintos centros de alta densidad. Esto sucede cuando la cobertura estudiada muestra áreas mal iluminadas, que acercan las nubes de puntos hacia el



Aunque en el plano *ik*, los agrupamientos *A* y *B* están alineados y ello pudiera hacer pensar de que se trata de una misma categoría, el examen de sus proyecciones sobre los planos *ij* y *jk* revela que se trata de dos agrupamientos diferentes. Su ángulo espectral será distinto de cero.

origen, frente a otras correctamente iluminados. Un efecto similar produce el exceso de humedad, que siempre oscurece la clase en las bandas del espectro visible y de los distintos infrarrojos.

La distancia de Mahalanobis proporciona un método para comparar distancias dentro de la métrica generada por cada agrupamiento, derivada de la distribución y densidad de los puntos de la nube. Se expresa así:

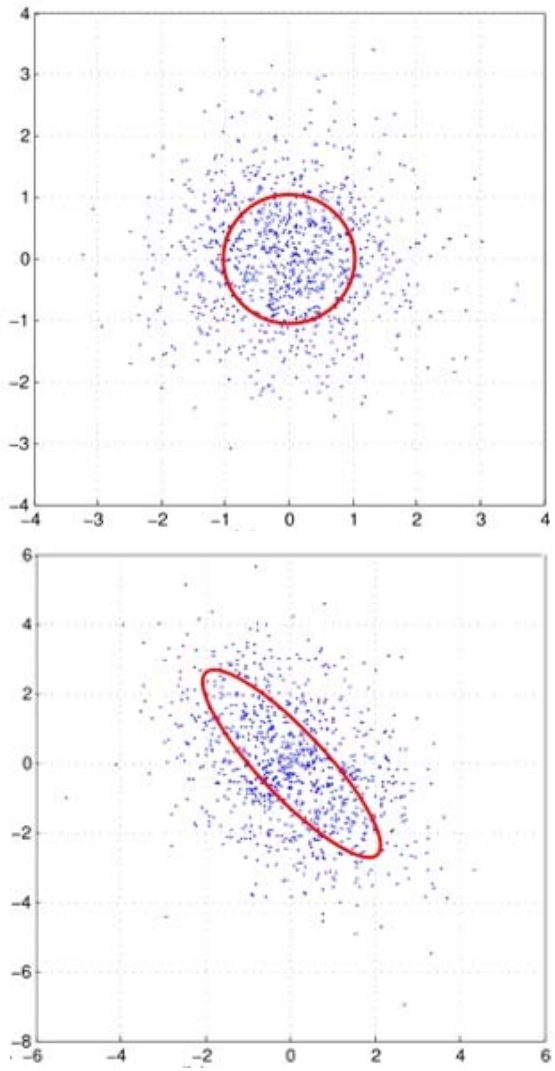
$$d_m\left(\vec{z},\vec{\mathrm{M}}\right)=\sqrt{\left(\vec{z}-\right)^t\Sigma^{-1}\left(\vec{z}-\vec{\mathrm{M}}\right)} \qquad (12)$$

y como puede observarse es realmente la raíz de la forma cuadrática de la distribución normal multivariante. En una nube hiperelipsoidal, las superficies de isodistancia no son hiperesferas sino hiperelipsoides, que quedan mejor caracterizados con la distancia de Mahalanobis.

Finalmente, otra distancia que suele emplearse con fines de separabilidad entre clases es la de Bhattacharyya:

$$D_B=\tfrac{1}{8}\left(\vec{\mathrm{M}}_2-\vec{\mathrm{M}}_1\right)^t\left(\frac{\Sigma_1+\Sigma_2}{2}\right)^{-1}\left(\vec{\mathrm{M}}_2-\vec{\mathrm{M}}_1\right)+ \\ +\tfrac{1}{2}\ln\frac{\left|\frac{\Sigma_1+\Sigma_2}{2}\right|}{\sqrt{\left|\Sigma_1\right|\left|\Sigma_2\right|}} \qquad (13)$$

Se considera un estadístico más fiable que la distancia de Mahalanobis, dado que esta última es un caso particular de la de Bhattacharyya cuando las desviaciones estándar de las dos clases son iguales. Cuando dos clases tienen valores promedio muy similares pero diferentes desviaciones estándar, la distancia de Mahalanobis tiende a cero, mientras que la distancia Bhattacharyya crece a medida que lo hace la diferencia entre las desviaciones estándar.



Uso de la distancia euclídea (arriba) y de Mahalanobis (abajo). Las líneas rojas representan los puntos de igual distancia al centro del agrupamiento.

Fase de asignación

Durante la fase de asignación se asigna una etiqueta de clase a cada celda, es decir, se atribuye cada celda de la imagen a una de las categorías de la leyenda previamente elaborada.

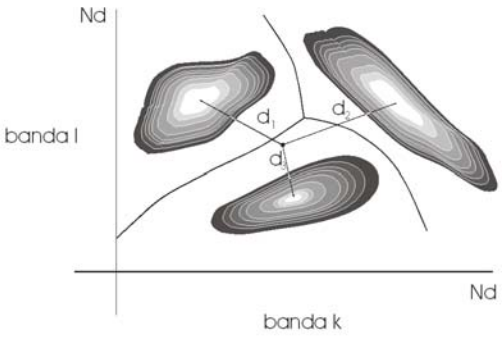
Para ello es necesario establecer una función de similaridad que permita asegurar que las celdas que pertenecen a un agrupamiento son más parecidas entre sí que a cualesquiera otras que pertenezcan a otro diferente. Las funciones de similaridad más habituales son las siguientes:

Distancia mínima: $d_{\min}\left(Z_i,Z_j\right)=\min\limits_{\substack{z\in Z_i\\ z'\in Z_j}}\left\|\vec{z}-\vec{z}'\right\|$ (14)

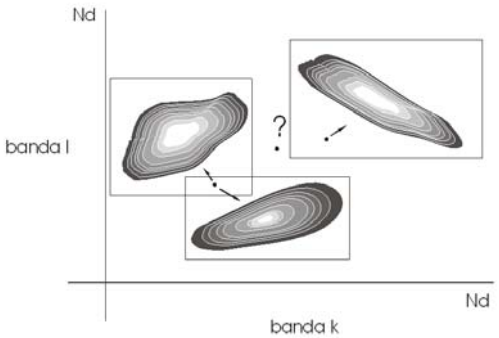
Distancia máxima: $d_{\max}\left(Z_i,Z_j\right)=\max\limits_{\substack{z\in Z_i\\ z'\in Z_j}}\left\|\vec{z}-\vec{z}'\right\|$ (15)

Distancia media: $d_{med}\left(Z_i,Z_j\right)=\left\|\vec{\mathrm{M}}_i-\vec{\mathrm{M}}_j\right\|$ (16)

Y en todas ellas, la norma $\|X\|$ utilizada es la euclídea:



Criterio de asignación por mínima distancia.



Criterio de asignación por paralelepípedos, aplicado a un espacio espectral de sólo dos dimensiones.

$$\|X\| = \sqrt{\sum_{i=1}^{n} |X_i|^2}$$

Clasificadores paramétricos

Los clasificadores son los operadores encargados de etiquetar los vectores de características $\vec{z}$ con alguna de las clases w_i comparando una conjunto de c funciones discriminantes $g_i(\vec{z})$, donde i=1, 2,... c:

$$\vec{z} \in w_i \qquad \text{si} \qquad g_i(\vec{z}) > g_j(\vec{z}) \qquad \forall\ j \neq i \tag{17}$$

El clasificador divide el espacio de características en c clases diferentes o regiones de decisión $w_1, w_2, \ldots w_c$ y asigna la etiqueta de clase w_i a un determinado vector de características $\vec{z}$ cuando la función discriminante que le corresponde para esa clase supera en valor a cualquiera del resto de las clases.

Entre los clasificadores más empleados se encuentran el de mínima distancia, el paralelepipédico, el isodata, el de k-Means, y el de máxima probabilidad.

- **Clasificador de mínima distancia**. El criterio más sencillo para asignar una celda a una categoría consiste en incluirlo en la espectralmente más cercana, dentro del espacio de características. La función de similaridad utilizada es la distancia euclídea entre el vector de características de la celda y el centro de la clase, representado por su vector de medias:

$$d_C = \sqrt{\sum_{i=1}^{n} \left(ND_i - \overline{ND}_{Ci} \right)^2} \tag{18}$$

La celda se asignará a la clase que presente una menor distancia d_C.

- **Clasificador paralelepipédico**. La clasificación paralelepipédica o de hipercubos fija áreas de dominio para cada categoría. Una celda de nivel digital $ND^{(k}$ en la banda k de una imagen de n bandas será atribuida a la clase C si se verifica:

$$\overline{ND}_C^{(k} - R_C^{(k} < ND^{(k} < \overline{ND}_C^{(k} + R_C^{(k} \qquad \forall 1 \leq k \leq n \tag{19}$$

siendo siendo $R_C^{(k}$ un rango de variación concedido alrededor del valor medio de la clase C en la banda k. Cuanto mayor sea el rango, mayor será el número de celdas incluidas en la categoría, pero se incrementará el riesgo de error de comisión de la clasificación. Por el contrario, si el umbral es excesivamente pequeño, ese tipo de error se minimizará, pero también aumentarán las celdas sin clasificar, incrementándose así el error de omisión.

- **Clasificador isodata**. Es un algoritmo de clasificación automática iterativo basado en la función de mínima distancia euclídea. En cada ciclo de clasificación se asignan las celdas a una clase provisional. Durante el desarrollo del algoritmo se registran fugas de celdas de una clase a otra hasta que el proceso converge. El algoritmo de clasificación isodata sigue los siguientes pasos:

1. Se eligen los valores iniciales $\vec{m}_1, \vec{m}_2, \cdots, \vec{m}_c$ que actuarán como semillas de los *clusters*. Si no hay valores previos, se toma el vector de características de la primera celda de la imagen.
2. Calcular la distancia entre la semilla y la segunda celda.

   a. Si es inferior a un radio espectral *r* prefijado, se incluye en el mismo *cluster*.
   b. En caso contrario se considera la segunda celda como segunda semilla, y así sucesivamente hasta *c* clases.
3. Se clasifica la imagen completa, asignando las celdas al agrupamiento más próximo.
4. Se recalculan las medias de los agrupamientos considerando las nuevas asignaciones, pasando a tomar los valores $\vec{m}_1', \vec{m}_2', \cdots, \vec{m}_c'$.
5. Se reitera el ciclo de reclasificación hasta mejorar la convergencia de los valores, o hasta alcanzar un número prefijado de iteraciones.
6. Se descarta cualquier agrupamiento que contenga menos datos que un determinado valor también prefijado.
7. Se reagrupan o se escinden los agrupamientos, de acuerdo con las siguientes reglas:
   a. Agrupar *clusters* si su número excediese de un valor prefijado.
   b. Dividir los *clusters* si su número no alcanza un valor umbral.
8. Se toma el centro de masas cada *cluster* como una nueva semilla y se reitera la clasificación como en la etapa 2.
9. Se repiten los pasos 3, 4, 5, 6 y 7 hasta que el proceso converja o hasta alcanzar un número de iteraciones establecido con anterioridad.

¬ **Clasificador k-*means*.** Este algoritmo de clasificación automática opera de un modo parecido al anterior, pero si en aquél se establecían las regiones de decisión como círculos en un plano euclídeo espectral R^2, en éste se disponen como polígonos de Thiessen. El algoritmo sigue los siguientes pasos:

1. Se especifica un número inicial de clases y se generan las clases por distancia euclídea. Se toma una celda de cada clase como centroide del agrupamiento. Se definen los segmentos que unen todos los centroides entre sí y se trazan sus mediatrices, con las que se generarán tantos polígonos de Thiessen como clases se han establecido.
2. Se configuran las nuevas clases con las celdas que pertenecen a cada polígono. Es posible que celdas que anteriormente pertenecían a una categoría, tras la teselación del espacio espectral, ahora pertenezcan a otra.
3. Se calcula el centro de masas de las nuevas nubes de puntos, es decir, sus vectores de medias, y se erigen como nuevas semillas de los agrupamientos para la segunda iteración.
4. Se reiteran los pasos 2,3 y 4 hasta que el proceso sea convergente y no se registren más fugas de celdas entre categorías, o su número sea inferior a uno establecido con anterioridad.

Naturalmente, tanto este clasificador como el isodata, trabajarán con la dimensionalidad completa de la imagen en un espacio generalizado, y distribuirán las regiones de decisión en hiperplanos. R^n, siendo *n* en el número de bandas.

¬ **Clasificador de máxima probabilidad.** Este clasificador es de los más robustos y utilizados dentro de los procedimientos paramétricos de clasificación digital. Consiste en etiquetar las celdas con la categoría a la que más probabilidad tenga de pertenecer por su vector de características, asumiendo distribuciones normales para las clases. Para ello se establece la probabilidad de pertenencia a cada clase, utilizando la función de densidad de probabilidad enunciada en (2).

Clasificador de máxima probabilidad bayesiana. La aplicación del teorema de Bayes a la clasifica-

ción de máxima probabilidad mejora los resultados finales. La probabilidad de un determinado suceso *B* condicionada a otro suceso *A* es la probabilidad de que ocurra B, supuesto que también se verifica la ocurrencia de A: $P(B/A)=\dfrac{P(A\cap B)}{P(A)}$ (20)

De igual forma:

$$P(A/B)=\frac{P(A\cap B)}{P(B)} \tag{21}$$

de modo que se puede escribir:

$$P(A\cap B)=P(A)P(B/A)=P(B)P(A/B) \tag{22}$$

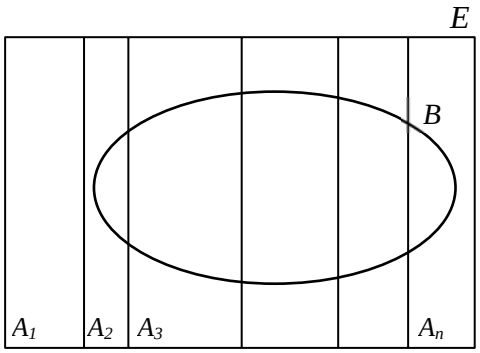
Dos sucesos se consideran independientes cuando:

$$P(A\cap B)=P(A)P(B) \tag{23}$$

Lo cual quiere decir que $P(A/B)=P(A)$ y que $P(B/A)=P(B)$. De lo contrario, son dependientes entre sí.

Pues bien, si en un contexto de decisión se consideran *n* sucesos mutuamente excluyentes A_i que formen un sistema exhaustivo, es decir que la unión de todos ellos dé lugar al suceso seguro o universo muestral *E*:

$$\bigcup_{i=1}^{n} A_i = E \tag{24}$$



Universo muestral *E* teselado por *n* sucesos A_i y, sobre ellos, el suceso *B*. Las probabilidades condicionadas $P(B/A_i)$ son las proporciones de elipse que interceptan cada rectángulo A_i en relación con la superficie de este último.

y de todos ellos se conocen las probabilidades $P(A_i)$ de que sucedan, la probabilidad del suceso *B*, del que se conocen todas las probabilidades condicionadas $P(B/A_i)$, será:

$$P(B)=P(E\cap B)=P\left[\left(\bigcup_{i=1}^{n}A_i\right)\cap B\right]=\sum_{i=1}^{n}P(A_i\cap B) \tag{25}$$

o sea:

$$P(B)=\sum_{i=1}^{n}P(A_i)P(B/A_i) \tag{26}$$

La expresión (22), aplicada a este caso, se puede poner:

$$P(A_i\cap B)=P(A_i)P(B/A_i)=P(B)P(A_i/B) \tag{27}$$

o sea:

$$P(A_i/B)=\frac{P(A_i)P(B/A_i)}{P(B)} \tag{28}$$

Sustituyendo el denominador por su valor (26):

$$P(A_i/B)=\frac{P(A_i)P(B/A_i)}{\sum_{i=1}^{n}P(A_i)P(B/A_i)} \tag{29}$$

que es la expresión del teorema de Bayes. En esta expresión, $P(A_i/B)$ es la probabilidad *a posteriori*, $P(A_i)$ es la probabilidad *a priori* y

$P(B/A_i)$ se denomina *verosimilitud*.

En el contexto de decisión dentro de la clasificación digital de imágenes en teledetección, la variable aleatoria es el vector de características de una celda $\vec{z}$, y el teorema de Bayes puede escribirse así:

$$P(w_i/\vec{z}) = \frac{P(w_i)P(\vec{z}/w_i)}{\sum_{j=1}^{s} P(w_j)P(\vec{z}/w_j)} \qquad (30)$$

Si w_1 y w_2 son dos posibles estados para una celda, es decir, dos clases a las que hay probabilidad de que pertenezcan, se asume la ocurrencia del suceso w_1 cuando:

$$P(w_1/\vec{z}) > P(w_2/\vec{z}) \qquad (31)$$

Utilizando el clasificador bayesiano, dado que los denominadores de la expresión (30) son siempre los mismos, sea cual sea la clase considerada, la expresión (31) conduce a la siguiente operación de decisión:

$$\text{decidir} \begin{cases} w_1 \text{ si } P(w_1)P(\vec{z}/w_1) \geq P(w_2)P(\vec{z}/w_2) \\ w_2 \text{ si } P(w_1)P(\vec{z}/w_1) < P(w_2)P(\vec{z}/w_2) \end{cases} \qquad (31)$$

en cuya expresión se han utilizado solamente los numeradores de la expresión de Bayes como función de decisión.

De modo general, la función discriminante en la decisión bayesiana es la probabilidad *a posteriori*:

$$g_i(\vec{z}) = P(w_i/\vec{z}) \qquad (32)$$

y la celda será atribuida a la clase para la que la probabilidad a posteriori sea más alta. Aplicando el teorema de Bayes:

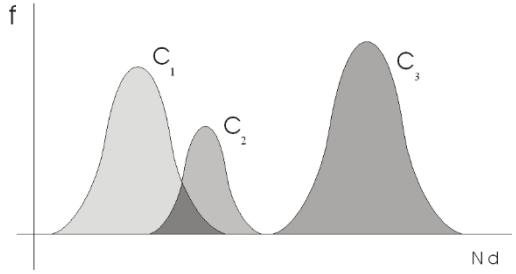
$$g_i(\vec{z}) = P(\vec{z}/w_i)P(w_i) \qquad (33)$$

De modo que la celda pertenece a la clase w_i si $g_i(\vec{z}) > g_j(\vec{z})$, o lo que es lo mismo:

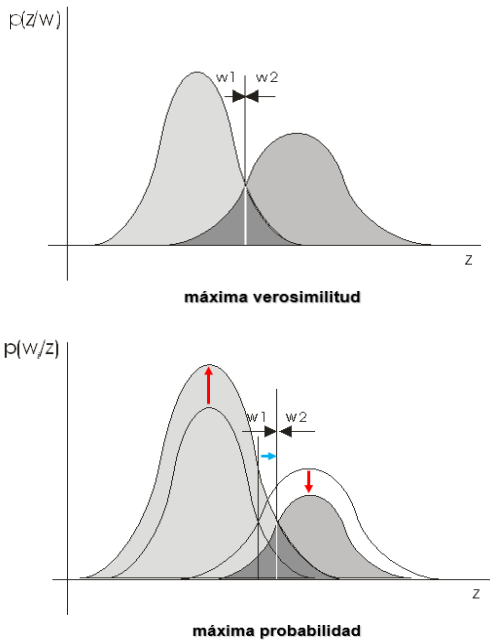
$$\text{decidir } w_i \text{ si } P(w_i/\vec{z}) > P(w_j/\vec{z})$$

$$\text{es decir, si } \quad P(w_i)P(\vec{z}/w_i) > P(w_j)P(\vec{z}/w_j) \qquad (32)$$

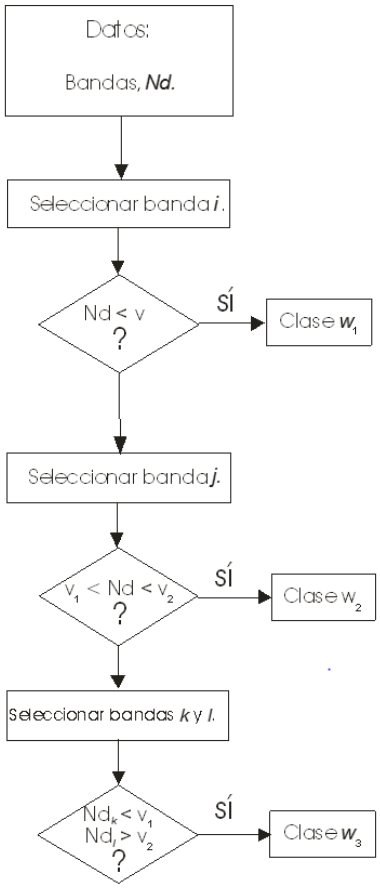
La función de decisión es, pues, la probabilidad *a posteriori* $P(w_i/\vec{z})$, pero no es fácil decidir a qué clase puede pertenecer una celda condicionada a que se verifiquen unos determinados niveles digitales, o lo que es lo mismo, a la vista de su vector de características. En cambio, resulta más fácil plantear el problema de la decisión en términos bayesianos. Por un lado, se conoce, o se puede conocer, la probabilidad *a priori* de cada clase $P(w_i)$, es decir, su abundancia dentro de la imagen que se va a clasificar o, lo que es lo mismo, la proporción que ocupa cada clase dentro de la escena, sea cual fuere la ubicación de las celdas que le corresponden Esta información puede obtenerse a partir de clasificaciones anteriores, u otras fuentes auxiliares. En el caso de no contar con dicha información, se puede adoptar provisionalmente una probabilidad igual a todas las categorías para iniciar la clasificación.



Un clasificador de máxima verosimilitud clasificará las categorías C_1, C_2 y C_3 como diferentes, aunque con errores de asignación entre la clase C_1 y C_2, debido al solapamiento de sus distribuciones.



El clasificador de máxima probabilidad baye-



Árbol de clasificación.

El otro factor pendiente de conocer para aplicar el teorema de Bayes es la verosimilitud $P(\vec{z}/w_i)$, que en este contexto es la probabilidad de que se cumpla un determinado nivel digital, o un vector de características si se habla de una imagen multiespectral, condicionado a cada clase de la leyenda. Si se asume que los niveles digitales de las clases siguen distribuciones normales multivariantes, la verosimilitud es simplemente la probabilidad leída en el eje de ordenadas de la gráfica de la distribución normal, o en las tablas estadísticas habituales.

Otros criterios de asignación

Hay otros clasificadores que no se atienen a los parámetros de la distribución normal. Entre ellos pueden citarse los siguientes.

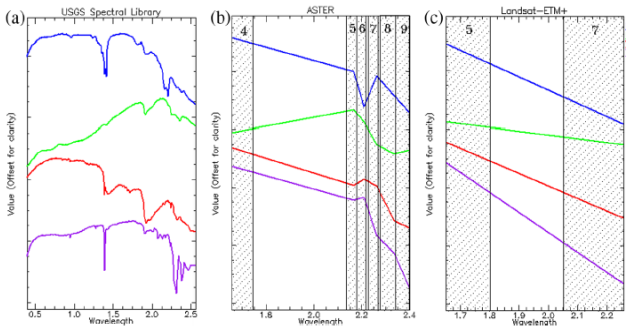
- **Clasificador en árbol.** Estos operadores identifican las categorías secuencialmente y en cada paso utilizan las bandas más apropiadas para la discriminación. Para configurar una determinada clase se seleccionan las celdas cuyo ND esté comprendido dentro de un cierto intervalo, o bien aquellas que cumplan una determinada condición, y se continúa el proceso sólo con ellas. Cuando, en una determinada banda, una celda exhiba un ND con el que pueda ser asignada a más de una categoría, el clasificador establecerá nuevas condiciones que involucren a varias bandas simultáneamente.

  Estos clasificadores suelen estar constituidos por centenares de reglas y proporcionan resultados muy satisfactorios, siempre la base de conocimiento sea suficiente. En ellos se basan los sistemas expertos que, además, aprenden de los resultados previos para sucesivas clasificaciones.

- **Clasificadores texturales.** En análisis de imagen, la textura se refiere a las diferencias de reflectancia entre celdas adyacentes. Estos clasificadores utilizan determinadas características en la vecindad de un pixel, como la reflectancia media, la varianza, la autocorrelación, etc., con el fin de incorporar a la clasificación algunas de las propiedades cualitativas que suelen ser tenidas en cuenta en fotointerpretación. Puede realizarse un análisis de la textura de una escena aplicando determinados filtros e incorporando esta información como una fuente más de discriminación.

- **Clasificadores de contexto.** Todos los clasificadores enunciados hasta ahora clasifican la imagen celda a celda utilizando cualquiera de los criterios explicados. Sin embargo, en las imágenes reales, lo más habitual es que las clases se presenten en la escena en grupos de celdas más o menos grandes, y con mucha menor frecuencia como celdas aisladas. Los clasificadores de contexto clasifican las celdas por grupos, previamente reunidos por similitud de los ND, etiquetando cada una de ellas en relación con la asignación de las adyacentes.

- **Clasificación hiperespectral.** Además de las técnicas clásicas de clasificación paramétrica, la teledetección hiperespectral utiliza procedimientos específicos que aseguran una mayor discernibilidad entre categorías, pudiéndose con ello separar eficazmente docenas de clases. En ocasiones no se requiere clasificar completamente la escena sino solamente identificar la presencia de ciertas clases a lo largo de ella. Para ello se realizan cocientes y otras operaciones algebraicas con las bandas, basados en rasgos espectrales únicos y específicos de los componentes buscados.



Sucesiva degradación de la signatura espectral: la obtenida mediante un espectrorradiómetro (a), la obtenida por el sensor Aster (b) y la de Landsat-ETM+ (c) de cuatro minerales de la arcilla: montmorillonita (azul) nontronita (verde), vermiculita (rojo) y talco (magenta). Fatima, K. *et al.* 2017. Minerals identification and mapping using satellite image. *Journal of Applied Remote Sensing*. SPIED.

De modo general, siempre es preferible un mayor número de bandas para establecer un mayor número de clases, tanto si se trabaja con operadores paramétricos como si se emplean otros procedimientos. La signatura espectral queda mejor establecida realizando barridos espectrales minuciosos que con los diagramas de signaturas con los que trabajan los operadores multiespectrales clásicos.

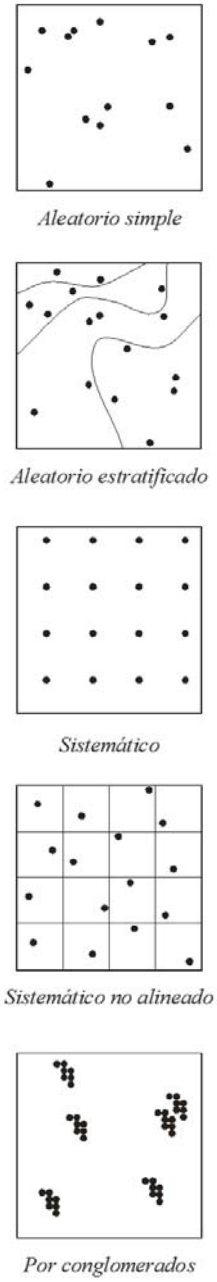
Control de calidad temático

La imagen clasificada no tiene una fiabilidad absoluta. Existen errores de asignación por similitud espectral entre celdas que, perteneciendo a categorías cartográficas diferentes, presentan signaturas espectrales muy parecidas.

Aunque los errores de asignación se refieren siempre a un mal etiquetado de las celdas dentro de una categoría de clase determinada, en el control de calidad estas incorrecciones pueden estudiarse desde dos puntos de vista, que dan lugar a dos tipos de error temático en el proceso de clasificación:

- **Error de omisión**. Se produce cuando, perteneciendo la celda a una determinada categoría, no ha sido asignada a ella. Es el denominado en estadística error tipo I.
- **Error de comisión**. Se produce cuando la celda es etiquetada dentro de una categoría a la que realmente no pertenece. Es el error tipo II.

Con el fin de ofrecer al usuario la información completa del proceso de clasificación, es necesario realizar un control de calidad que establezca la fiabilidad del proceso. Para ello, tras la clasificación, se debe realizar un proceso de verificación. Si se dispone de una fuente de verdad-terreno exhaustiva, por ejemplo, procedente de fuentes cartográficas o de otras clasificaciones anteriores, la verificación puede abarcar la imagen completa. Sin embargo, lo más frecuente es no disponer de esa información, por lo cual



Tipos de muestreo de verificación.

es necesario plantear una campaña de muestreo para la verificación.

Existen varios tipos de muestreo de verificación. Entre los más difundidas están los siguientes.

- **Aleatorio simple**: las celdas a verificar se designan al azar.
- **Aleatorio estratificado**: se divide el espacio muestral en diversas subpoblaciones o estratos sobre cada una de los cuales se aplica un muestreo aleatorio.
- **Sistemático**: a partir de un punto designado al azar, se confecciona la muestra mediante la selección de puntos equidistantes en las dos dimensiones de la imagen.
- **Sistemático no alineado**: se traza una cuadrícula sistemática, a partir de cuyos nudos se varían al azar las coordenadas de los elementos de la muestra.
- **Por conglomerados**: es un muestreo aleatorio en el que en lugar de una sola celda se selecciona un grupo de ellas para formar la muestra, siguiendo siempre el mismo patrón geométrico.

Lo que la verificación de la exactitud de la clasificación trata de establecer es el nivel de error en el etiquetado de las celdas. Para ello, juega un papel fundamental el tamaño de la muestra. El número de celdas que deben verificarse depende de dos factores. En primer lugar, del nivel de confianza que quiera otorgarse a la estimación. Es obvio que para tener el 100 % de confianza en la cifra de error debería verificarse el conjunto completo de celdas de la imagen. En segundo lugar, de la información previa que se posea sobre la población. El conocimiento previo de la dispersión de ND dentro de cada clase, es decir, su varianza, permite reducir el número de muestras sin disminuir el nivel de confianza de la estimación.

El suceso {celda bien clasificada-celda erróneamente clasificada} se distribuye como una variable aleatoria discreta binomial:

$$X \in b(n, p) \tag{33}$$

donde n es el tamaño de la muestra y p es la probabilidad de acierto en la clasificación. La función de densidad de probabilidad de la distribución binomial es:

$$f(x) = P(X=x) = \binom{n}{x} p^x q^{n-x} \tag{34}$$

con $x \in [0,1,\cdots,n]$ y $q=1-p$, es decir, la probabilidad de error en la clasificación.

Según el teorema de Moivre, cuando $n \to \infty$, la distribución binomial es asintóticamente normal $N(np, npq)$, es decir, converge en distribución a una normal de $\mu = np$ y $\sigma^2 = npq$. En la práctica, el criterio de convergencia puede adoptarse cuando $n \geq 30$.

Con todo ello, el problema de determinar el tamaño muestral se plantea así: se trata de calcular el tamaño n de la muestra para estimar la probabilidad p de que una categoría esté bien clasificada, con un error de $\pm\varepsilon$ y un nivel de confianza en la estimación de $1-\alpha$, esto es, que el error tipo I (rechazar la hipótesis H_0 de asignación acertada, siendo ésta verdadera) sea inferior a α. Se trata, pues, de un problema de estimación por intervalo de confianza, donde éste es:

$$I_{1-\alpha} = \left[\hat{p} - z_{\alpha/2}\sqrt{\frac{pq}{n}},\ \hat{p} + z_{\alpha/2}\sqrt{\frac{pq}{n}} \right] \tag{35}$$

siendo $z_{\alpha/2}$ el percentil de la distribución normal $N(0,1)$ correspondiente a la mitad del nivel de confianza exigido en la estimación.

El error admitido en la asignación de etiquetas a la celda será:

$$\varepsilon = z_{\alpha/2}\sqrt{\frac{pq}{n}} \tag{36}$$

De donde puede deducirse el mínimo número de celdas que deben integrar la muestra de verificación:

$$n = \frac{z_{\alpha/2}^2 pq}{\varepsilon^2} \tag{37}$$

Si no se dispone de información más precisa sobre la población, en el peor de los casos puede adoptarse $p = q = 0{,}5$, de tal modo que pueda disponerse de un límite inferior para el tamaño muestral:

$$n \geq \frac{z_{\alpha/2}^2}{4\varepsilon^2} \tag{38}$$

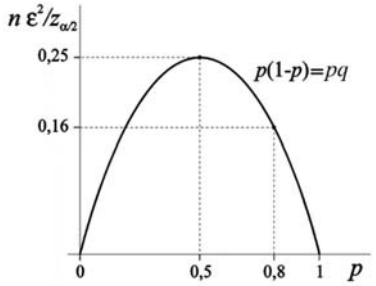
Este valor es válido para determinar el número de celdas a muestrear para cada categoría. El número total de celdas de verificación será el producto de este valor por el número de clases de la leyenda. Si, por el contrario, se conoce *a priori* el intervalo de existencia de la probabilidad p de resultar la celda bien clasificada, se puede reducir el tamaño de la muestra, en la medida que se reduce el producto pq.

Es bastante frecuente verificar los resultados a partir de un muestreo sistemático no alineado o, en su defecto, uno aleatorio simple. Para cualquiera de estos casos, si se requiere un nivel de confianza del 99 %, el percentil buscado es $1\text{-}\alpha = 0{,}99$; $\alpha = 0{,}01$ y $\alpha/2 = 0{,}005$. El valor de z que proporciona ese nivel de confianza es:

$$z_{\alpha/2} = 2{,}575; \quad z_{\alpha/2}^2 = 6{,}630625$$

Admitiendo un error máximo del ±5 % en la asignación de la celda, el tamaño de la muestra de verificación deberá ser como mínimo:

$$n = \frac{z_{\alpha/2}^2}{4\varepsilon^2}; \; n = \frac{6{,}630625}{4 \cdot 0{,}05^2} = \frac{6{,}630625}{0{,}01} \approx 663$$



El peor escenario para determinar el mínimo número de celdas de la muestra de verificación es en el que no se conoce la probabilidad de ser bien clasificada la categoría. En ese caso, $p = q = 0{,}5$ y n es máximo.

Matriz de confusión

Los resultados de la verificación suelen presentarse en una tabla de contingencia cuadrada en la que las filas recogen las categorías obtenidas en el proceso de clasificación y las columnas la verdad-terreno, es decir, las categorías verificadas.

Cada elemento de la matriz representa el número de celdas que, perteneciendo en la clasificación a la categoría que marca su fila, realmente la verificación ha demostrado su pertenencia a la clase que indica su columna. La diagonal principal de la matriz representa el número de celdas correctamente clasificadas para cada categoría de la leyenda y los elementos fuera de la diagonal principal indican errores de asignación, que, como ha quedado establecido, son de dos tipos:

- Los situados dentro de una misma columna representan el error de omisión.
- Los pertenecientes a una misma fila representan el error de comisión.

		CATEGORÍAS VERIFICADAS												
		Agua 1	Suelos desnudos 2	Coníferas 3	Frondosas 4	Herbáceos extensivos 5	Regadío 6	Pradera 7	Monte bajo 8	Urbano cerrado 9	Urbano abierto 10	SUMA	Error absoluto de comisión	Riesgo del usuario (%)
CATEGORÍAS EN LA IMAGEN CLASIFICADA	Agua 1	3										3	0	0
	Suelos desnudos 2		2						1			3	1	33
	Coníferas 3			27				3	2			32	5	16
	Frondosas 4			1	7	1	2					11	4	36
	Herbáceos extensivos 5					86		1	4			91	5	5
	Regadío 6				3		19					22	3	15
	Pradera 7					1		9	1			11	2	18
	Monte bajo 8		2			1		2	31			36	5	24
	Urbano cerrado 9									9		9	0	0
	Urbano abierto 10									1	2	3	1	33
SUMA		3	4	28	10	89	21	15	39	10	2	221	26	
Error absoluto de omisión		0	2	1	3	3	2	6	8	1	0	26		
Riesgo del productor (%)		0	50	4	30	3	10	40	21	10	0			

Un ejemplo de matriz de confusión para 10 categorías.

El error de omisión, también llamado riesgo del productor, se calcula como la proporción de los residuos por columnas en relación con el total:

$$EO_j = \frac{\sum_i x_{ij} - x_{jj}}{\sum_i x_{ij}} \qquad i \neq j \tag{39}$$

Por su parte, el error de comisión, o riesgo del usuario, se calcula como la proporción de los residuos por filas en relación con el total:

$$EC_i = \frac{\sum_j x_{ij} - x_{ii}}{\sum_j x_{ij}} \qquad i \neq j \tag{40}$$

Los complementarios de los errores de omisión y comisión se llaman fiabilidades del productor y del usuario, respectivamente:

$$FP_j = \frac{x_{jj}}{\sum_i x_{ij}} \tag{41}$$

$$FU_i = \frac{x_{ii}}{\sum_j x_{ij}} \tag{42}$$

Cuando la clasificación solamente tiene dos categorías, resulta evidente comprobar que el error de omisión de una de ellas es justamente el de comisión de la otra. Sin embargo, cuando el número de clases es mayor, los errores se fugan entre las categorías y solamente se puede afirmar que la suma de los errores de omisión es igual a la suma de los de comisión, lo cual puede comprobarse en los ejemplos al margen.

Dentro del ámbito del análisis categórico multivariante, uno de los índices más empleados en la evaluación de la calidad de la clasificación es el porcentaje de acuerdo, que se refiere a la concordancia observada, esto es, la suma de los elementos situados en la diagonal de la matriz de confusión, que son los elementos correctamente identificados en la clasificación, en relación con el total de la muestra, expresada en tanto por ciento:

$$P_a = \frac{\sum_i x_{ii}}{n} \cdot 100 \qquad (43)$$

Y otro índice muy empleado en la evaluación de la calidad de la clasificación es el estadístico *Kappa* κ de Cohen, que cuantifica el nivel de acierto atribuible al método de clasificación seguido, por encima del que se hubiese obtenido meramente por azar:

$$\kappa = \frac{P_a - P_e}{1 - P_e} \qquad (44)$$

donde P_a es la concordancia observada (porcentaje de acuerdo) y P_e la concordancia esperada por azar.

Y teniendo en cuenta (100), y que:

$$P_e = \frac{\sum \left[\sum_i x_{ij} \sum_j x_{ij} \right]}{n^2} \cdot 100 \quad i \neq j \qquad (45)$$

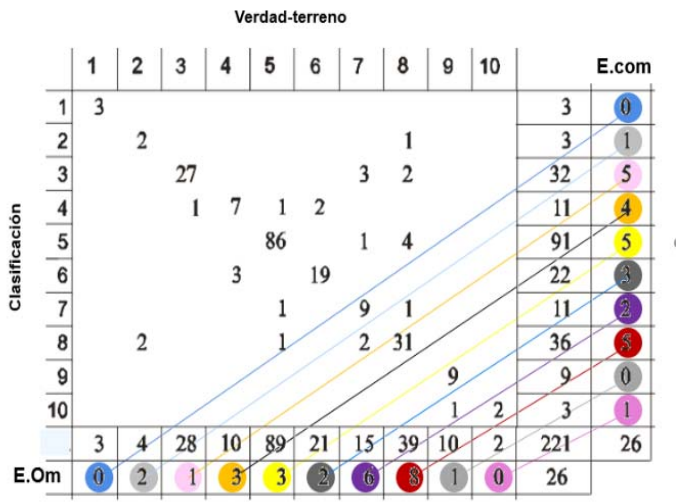
finalmente se tiene:

$$\kappa = \frac{n \sum_i x_{ii} - \sum \left[\sum_i x_{ij} \sum_j x_{ij} \right]}{n^2 - \sum \left[\sum_i x_{ij} \sum_j x_{ij} \right]} \qquad (46)$$

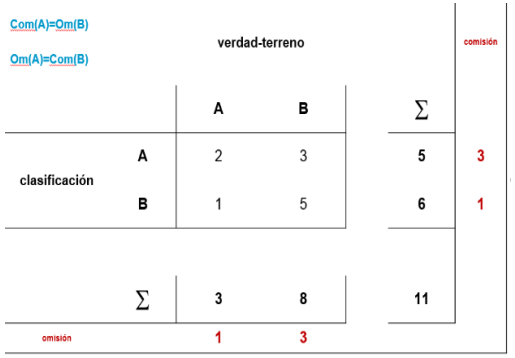
El índice κ puede adoptar valores comprendidos entre 0 y 1. La interpretación de los resultados puede seguir el criterio recogido en la Tabla 1:

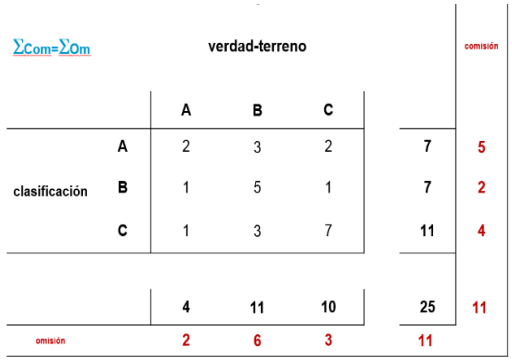
Tabla 1. Significado de los valores de Kappa.

Valor de κ	Concordancia
0,00	Pobre
0,01-0,20	Leve
0,21-0,40	Aceptable
0,41-0,60	Moderada
0,61-0,80	Considerable
0,81-1,00	Casi perfecta



Forma de calcular el estadístico κ.





Cuando el número de clases excede de dos, solo se puede asegurar que la suma de errores de omisión es igual a la de los de comisión